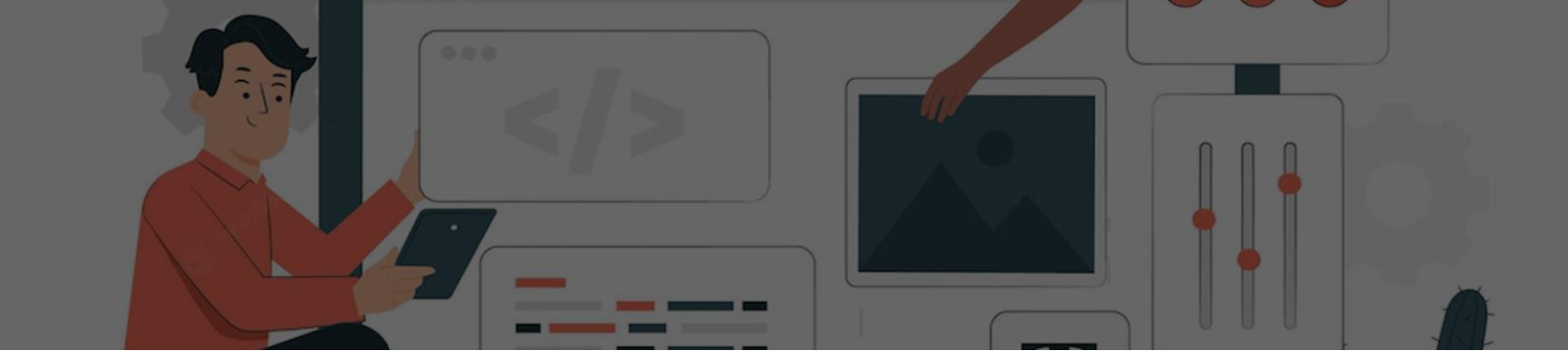

IMSA data science course

Date: 'Bahman' 6th – 'Esfand' 3rd

Project Manager: Arash Rahimi **Committee:** SCOME

Course Summary

This course is designed to provide participants with a comprehensive introduction to the principles and applications of data science in the medical field. Through a combination of programming, statistics, machine learning, and advanced topics like bioinformatics and medical image analysis, students will gain the skills necessary to analyze and interpret medical data effectively.

Course structure

Module 1: programming basics (6,7,13,14,20,21 of "Day")

1. **Introduction to Data Science in Medicine**
    a. Importance of data science in medical research and practice.
    b. Examples of impactful studies using data science.
    c. Overview of the course structure.
    d. Tools and software setup (Python, R, Jupyter Notebooks, etc.).

2. **Introduction to Programming for Data Science (Python-focused)**
    a. What is programming? Why is it important for medical research?
    b. Setting up your environment (Python, Jupyter Notebook, or Google Colab).
    c. Python basics:
        i. Variables and data types.
        ii. Input/output.
        iii. Simple calculations.

3. **Control Structures and Loops**
    a. Conditional statements (if, else, elif).
    b. Loops (for, while).

4. **Functions and Modular Programming**
    a. What are functions? Why use them?
    b. Writing reusable code with functions.

5. **Working with Lists, Dictionaries, and Strings**
    Data structures in Python:
        i. Lists and loops.
        ii. Dictionaries for structured data.
        iii. String manipulation.
6. **Introduction to Libraries for Data Science**
    Overview of Python libraries:
        i. **NumPy** for numerical operations.
        ii. **Pandas** for data manipulation.

Module 2: Statistics (27,28 of "Day", 4,5,11,12 of "Bahman")

7. **Basic concepts**
    a. Study designs and sampling types
    b. Describing a single variable: central tendency, index of dispersion
8. **Data visualization**
    a. Scatter, histogram, pie chart, etc
    b. Tools: matplotlib
9. **Distributions**
    a. Discrete distributions: binomial, Poisson, hypergeometric
    b. Continuous distributions: normal
10. **Inferential statistical**
    a. H0 /H1 , p-value, one/two-tailed, doing a test
    b. Z-test, t-test (all 3 types + short overview of degree of freedom), chi squere
    c. Tools: statsmodel
12. **Advanced Statistical Methods 2x**
    a. Linear, logistic, multiple regression models
    b. ANOVA and Survival analysis and multivariate analysis (PCA)

Module 3: Machine learning (18,19,25 of "Bahman")

13. **Data Collection and Cleaning**
    a. Sources of medical data (e.g., electronic health records, public databases).
    b. Handling missing data and outliers.
    c. Practical: Cleaning a real-world medical dataset.
14. **Supervised/unsupervised and model introduction**
    a. Overview of machine learning techniques (supervised vs. unsupervised).
    b. Common algorithms: Linear regression, logistic regression, and decision trees.
    c. Practical: Building a logistic regression model to predict disease outcomes.
15. **Evaluation Metrics**
    a. sensitivity, specificity, and ROC curves

Module 4: Extra (26 of "Bahman",2,3 of "Esfand")

16.Bioinformatics

a. Overview of bioinformatics tools.
b. Introduction to genomic data and sequence analysis.
c. Practical: Analyzing a small genomic dataset (e.g., using Biopython).

18. Introduction to Image Analysis in Medicine 2x

a. Understand the importance of image analysis in medical practice and research.
b. Perform basic medical image processing and segmentation tasks.
c. Use tools like LabelImg or Python-based annotations for segmentation.

Capstone project

Introduction to search and writing a proposal (Ms. Khan Mohammadi)

Objective: Apply knowledge from all modules to solve a real-world medical data science problem.

- Analyze a patient dataset to uncover insights (e.g., BMI trends, disease correlations).
- Develop a simple machine learning model to predict disease risk.